PNAS
www.pnas.org

Supplementary Information for

Protein Folding from Heterogeneous Unfolded State Revealed by Time-resolved X-ray Solution Scattering

Tae Wu Kim,[†,‡,§] Sang Jin Lee,[†,‡,§] Junbeom Jo,[†,‡,§] Jong Goo Kim,[†,‡,§] Hosung Ki,[†,‡,§] Chang Woo Kim,[†] Kwang Hyun Cho,[†] Jungkweon Choi,[§] Jae Hyuk Lee,[¶] Michael Wulff,[#] Young Min Rhee,[†,*] and Hyotcherl Ihee[†,‡,§,*]

[†]Department of Chemistry and [‡]KI for the BioCentury, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, 34141, Republic of Korea
[§]Center for Nanomaterials and Chemical Reactions, Institute for Basic Science (IBS), Daejeon, 34141, Republic of Korea
[¶]XFEL Beamline Department, Pohang Accelerator Laboratory, Pohang, 37673, Republic of Korea
[#]Experiments Division, European Synchrotron Radiation Facility, F-38043 Grenoble Cedex, France

Young Min Rhee and Hyotcherl Ihee
Email: Y.M.R., ymrhee@kaist.ac.kr
H.I., hyotcherl.ihee@kaist.ac.kr

This PDF file includes:

www.pnas.org/cgi/doi/10.1073/pnas.1913442117

Supplementary Information Text

Note 1. Structural analysis of X-ray scattering data based on ensemble optimization method

We employed the ensemble optimization method (EOM) based on molecular dynamics (MD) simulations in order to extract the structural parameters of the protein such as radius of gyration (R_g) and three-dimensional protein conformation from the static X-ray scattering curves. This approach consists of two steps: (i) sampling various protein conformations implemented by MD simulations and (ii) obtaining the ensemble of protein conformations that reproduces the experimental scattering curve. The second step is achieved by using the sampled structures from the MD simulations with the EOM analysis. Below, we will provide more details on these two steps.

(i) Generation of protein conformations from molecular dynamics simulations. MD simulations of equine heart cyt-c were performed starting from the X-ray crystallographic structure with the PDB entry 1HRC (1). Of course, thermal fluctuations from the crystal structure in solution are naturally reflected with the simulations. GROMACS 4.5.5 package (2) together with the Gromos96-43a1 force field (3) was employed in combination with the TIP3P water model (4). The system temperature was maintained with the velocity-rescale thermostat (5). The protein molecule was solvated in a periodic cubic box with the side length of 6 nm, containing explicit waters and GdnHCl molecules (6). To neutralize the charge in the box, some water molecules were replaced by sodium ions. A cutoff distance of 1.6 nm was used for calculating Lennard-Jones interactions and the particle mesh Ewald (PME) algorithm (7, 8) was used for treating long-range electrostatics. Prior to a production run, the system was energy minimized with the steep descent algorithm and was subsequently equilibrated for 20 ps at 293 K. This was followed by the actual production run over the duration of 1 ns, and the conformations were sampled at every 1 ps. These 1,000 protein conformations were considered as the structure pool of the reduced state.

We also performed unfolding simulations in order to analyze the experimental data collected from the oxidized cyt-c. With the microscopic reversibility (9), the folding pathway is the reverse of the unfolding pathway under the same condition. Performing unfolding simulations by starting from a well-defined folded structure could in principle be an appropriate strategy of sampling all possible protein conformations of the unfolded state at equilibrium, but directly simulating chemically denatured state at room temperature will be computationally intractable. To circumvent this issue, we expedited unfolding by employing a high temperature (800 K) in order to effectively access the vast conformational space in a feasibly reachable timescale (10, 11). To further reduce the simulation time, in this case, we switched to the implicit generalized Born solvent-accessible surface area (GB/SA) model (12). Because the Gromos96-43a1 force field does not define interaction parameters for metal ligation involving Fe in the heme group, some important interactions around it such as Met80(S)-Fe and His18(N)-Fe bonds are not described. During the high-temperature unfolding simulations, therefore, we employed a physical constraint between His18-N and heme-Fe atoms. This condition consequently allowed dissociation of the Met80(S)-Fe bond originating from thermal perturbation, while keeping the His18(N)-Fe bond intact. Earlier studies reported that unfolded cyt-c still maintains the His18(N)-Fe bond even in the denaturing condition with 3.5 M GdnHCl (13, 14), and they

justify the physical constraint between His18 and the Fe atom adopted in our unfolding simulation. However, we did not enforce mis-ligations such as Fe-His26(N) or Fe-His33(N). This was partly because it was somewhat tricky to artificially generate such mis-ligated conformations for initiating MD simulations. Thus, the unfolded structures generated from our MD simulations can possibly be more extended relative to the mis-ligated states. Nevertheless, the main purpose of the unfolding simulations was to generate diverse unfolded conformations, not to describe the sophisticated structure around the heme moiety in a statistically correct manner. Furthermore, since the current scattering data covering the q-region less than 0.5 $Å^{-1}$ cannot have the atomic-level sensitivity, the ensemble histogram as a function of R_g will not likely be highly affected by the existence of the mis-ligated conformations. In total, six independent trajectories were computed with the stochastic Langevin dynamics algorithm after setting the solvent collision rate to 0.5 ps^{-1}. The nonbonded interaction cutoff was 1.5 nm. Each trajectory was first randomized somewhat for 1 ns, followed by a production run. In the production run, we employed three different simulation times of 100 ns, 500 ns, and 1 μs and systematically investigated the convergence of simulation trajectories to assess the simulation time required to properly sample candidate structures of unfolded state.

We combined the snapshots from both the folding and the unfolding simulations to generate six separate structure pools at each simulation time. Each pool with 51,000 conformations in 1-μs-long simulation was used as an input to the EOM analysis. Each pool from 500-ns-long simulation contained 26,000 conformations and the pool from 100-ns-long simulation had 22,000 conformations. As shown in Fig. S2, the R_g distributions (referred to as conformational space maps from the structure pool) have a broad distribution covering from ~15 Å to ~38 Å in the structure pool from the 1-μs-long simulation. All the conformational spaces encompass the R_g values (24–29 Å) of unfolded cyt-c reported with earlier experiments (15, 16).

(ii) Ensemble optimization method (EOM) based on genetic algorithm. Each pool of protein structures, generated from the MD simulations, is considered as the candidate conformations to be included in the ensembles for the folded and the unfolded states. These folded and unfolded conformations were used to calculate the theoretical X-ray scattering curves using CRYSOL program (17). To maximize the intensity of X-rays, the quasi-monochromatic beam (3 % bandwidth), which has a broad and asymmetric spectrum as shown in Fig. S15, from the undulator fundamental was used in the experiment. The polychromaticity in the X-ray spectrum brings about a slight shift and damping in the scattered intensity, $S(q)$. To consider the effect of quasi-monochromatic beam in the scattering signal, we convoluted the energy spectrum of X-ray pulse with the theoretical scattering curves calculated from the MD-generated protein conformations according to the already established protocols (18, 19).

Assuming that all conformers exist with the equal probability, a potential solution to describe the experimental data is represented by an ensemble with M different conformers. The averaged scattering curve for protein conformations in an ensemble was compared with the experimental X-ray scattering data:

$$S_{avr}(q) = \frac{1}{M}\sum_{n=1}^{M} S_n(q)$$

where $S_{avr}(q)$ is the averaged scattering curve for an ensemble, $S_n(q)$ is the theoretical scattering curve for the *n*-th conformer, and *M* is the total number of conformers in an ensemble. Throughout this work, *M* = 20 was adopted. For the optimization, a genetic algorithm was used to rapidly search the optimal ensemble from the comparison between the theoretical (averaged scattering curve) and the experimental data (20, 21). In the first stage of the genetic algorithm, 50 ensembles (chromosomes) were created by randomly selecting 20 different conformations from the pool of protein structures. In each generation, these 50 chromosomes were submitted to mutation and crossing operations. In mutation, 50 % constituents of each chromosome were exchanged from the unchosen structures in the pool or from other chromosomes in the same generation. In crossing, 40 % constituents of two randomly selected chromosomes in the same generation were exchanged by maintaining the number of conformers in each chromosome. After these two genetic operations, the number of total ensembles was three times increased from 50 chromosomes to 150 chromosomes. For each chromosome, the averaged scattering curve, $S_{avr}(q)$, of individual theoretical scattering curves was compared with the experimental scattering to yield the parameter χ^2:

$$\chi^2 = \frac{1}{K-1}\sum_{j=1}^{K}\left[\frac{\alpha \cdot S_{avr}(q_j) - S_{exp}(q_j)}{\sigma(q_j)}\right]^2$$

where $S_{exp}(q)$ is the experimental scattering curve, $S_{avr}(q)$ is the averaged theoretical scattering curve for an ensemble (chromosome), α is the scaling factor between the theory and the experiment, *K* is the number of experimental q-points, and $\sigma(q)$ is the experimental standard deviation. Among the 150 ensembles generated from the mutation and the crossing, the 50 chromosomes, possessing the lowest χ^2 value, were selected and propagated to the next generation (evolution of chromosome). These processes of performing the mutation, the crossing, and the evolution were repeated over 5,000 generations. After the completion of the evolution, the chromosome that best fits the experimental data was finally collected to extract the optimal structural parameter of R_g and the three-dimensional protein conformations. For a pool of protein structures, the EOM analysis was repeated by using five different starting points. The resultant distributions of R_g were averaged with calculating the standard deviation. This was depicted by the distribution with error bars in Fig. 2A-C. To check the dependence of the final result on the structure pool itself from the EOM, we applied the analysis to the six different pools generated from the independent MD trajectories.

To check if the MD simulations reveal any specific native-like contacts, mis-ligated events, or other non-native interactions for both folded and unfolded states, we have generated the average contact map from the unfolded ensemble after the EOM adjustment as shown in Fig. S5. In close examination of the contact map, a weak non-native contact was observed in one of unfolded structures with a population of about 7% shown in Fig. S5c. However, we could not observe any significant long-range contact that persistently exists in the unfolded ensemble. There do exist some short range contacts, mostly within a residue-residue distance of 7 Å, but even these are found with low probabilities. This suggests that such contacts are only transient in nature and the lack of persistent contacts is the result of the random-coil nature of the unfolded cyt-c. This

characteristics is also consistent with the quite flat CD spectral feature at high GdnHCl concentrations shown in Fig. S6.

Note 2. Solvent heating experiment and elimination of solvent heating contribution from time-resolved X-ray scattering data

As described in the main text, the difference scattering signals measured at positive time delays after photo-exciting solute molecules contain not only contributions related to structural changes of the protein but also contributions arising from the heating of the solvent as a consequence of energy transferred from the excited molecules. To extract only the contributions from the structural change of protein molecules, the solvent heating contribution needs to be eliminated properly from the difference scattering curves. To do so, we implemented an additional time-resolved X-ray scattering experiment using the pristine buffer solution with 20 mM NADH and 3.5 M GdnHCl but without any protein. Upon photoexcitation, NADH undergoes both the formation of photoproduct and the recovery of excited state into the ground state that induces dissipation of excitation energy into the solvent. As shown in Fig. S8b, the difference scattering signal of buffer solution in the small angle region is small and featureless, whereas the signal in the wide-angle region (0.8 Å^{-1} $< q <$ 2.1 Å^{-1}) is dominant. The solvent heating response in the wide-angle region was well matched with the appropriately scaled difference scattering curve of cyt-c in the wide-angle region ($0.8 < q < 2.1$ Å^{-1}) shown in Fig. S9. Therefore, the solvent heating response, determined from the separate experiment of pristine buffer solution, can be used to eliminate the solvent heating contribution from time-resolved X-ray scattering curves of cyt-c according to the established protocols (18, 22-25). The heating-free difference scattering curves were obtained by subtracting the scaled solvent heating response from the difference scattering curves of cyt-c at various positive time delays shown in Fig. 4A.

Note 3. Global kinetic analysis for the time-resolved X-ray scattering data

We applied singular value decomposition (SVD) to the solvent heating-free experimental scattering curves to determine the kinetics of structural transitions of the protein (26). By the SVD analysis, the data matrix of time-resolved difference scattering curves, **A**, can be decomposed into left- and right singular vectors using the relationship of $\mathbf{A} = \mathbf{USV}^{\mathbf{T}}$, where **U** consists of the left singular vectors (LSVs) corresponding to time-independent q spectra, **V** consists of the right singular vectors (RSVs) corresponding to time-dependent amplitude change of **U**, and **S** contains the weights of the singular vectors. The LSVs contains the information on the scattering curves of structurally distinct transient intermediates while the RSVs represent the population dynamics of those transient intermediates. Thus, the SVD analysis allows us to obtain the number of structurally distinct species and the dynamics of each species regardless of the kinetic model. From the SVD analysis in the time range from 31.6 μs to 316 ms, we identified two significant singular vectors judged from the singular values (**S**) and the autocorrelation factors of the corresponding singular vectors. Fig. S10a shows the first four LSVs while Fig. S10b shows the first four RSVs scaled with singular values obtained from the SVD analysis. By globally fitting the two rSVs by a stretched exponential sharing common

relaxation times, $e^{-(t/\tau)^{\beta}}$, the time constant (τ) of 185 (± 75) ms and the β value of 0.70 (± 0.08) were determined. To examine the applicability of sum of exponentials to fit the RSVs, we applied bi-exponentials sharing common relaxation times and the time constants of 16.4 (± 5.1) ms and 193 (± 34) ms were determined. Based on the stretched exponential kinetics with the transition of 1st species → 2nd species, we implemented principal component analysis (PCA) whereby the experimental scattering curves were decomposed into two species-associated difference scattering curves (SADS) corresponding to the two intermediates as follows:

$$\Delta S_{theory}(q_i, t_j) = \sum_{k=1}^{2} \left[C_k(t_j) \right] \Delta S_{C_k}(q_i)$$

Here, $\Delta S_{theory}(q_i, t_i)$ is the theoretical difference scattering curve at given q and t values, $\Delta S_{C_k}(q_i)$ is the SADS corresponding to the k-th intermediate species at a given q value, $C_k(t_j)$ is the instantaneous population of the k-th intermediate at a given t value and can be calculated using the time constants obtained from the SVD analysis. After that, we minimized the discrepancy between the theoretical and experimental scattering curves using Nelder-Mead simplex algorithm (27). The whole analysis was performed by the home-built Matlab code. In addition to the stretched exponential kinetics, we implemented the PCA analysis employing two additional candidate models; (i) bifurcated kinetics including parallel pathways between two states, and (ii) sequential kinetics based on the consecutive transitions involving three states. As shown in Fig. S11a-c, the number of SADSs is determined by the number of states. Namely, two SADSs for the stretched exponential kinetics and the bifurcated kinetics, and three SADSs for the sequential kinetics are obtained, respectively. The comparison of SADSs from the stretched exponential kinetics and the bifurcated kinetics shows that two kinetic models are indistinguishable. Regarding the sequential kinetics model that needs one additional state relative to the stretched exponential or the bifurcated kinetics, the key question is whether the additional component is meaningful as a mathematically independent form from the other components. If the added component is indeed essential in describing the entire experimental data, it will not be possible to represent that component as a linear combination of the other components. This is simply because each SADS should have its own characteristic spectral shape. The result shown in Fig. S11e actually contradicts this aspect. Namely, by using a linear combination of the 1st and the 3rd SADSs, the 2nd one can almost perfectly be reconstructed. Thus, we can evaluate that the 2nd SADS is not an essential component. This means that assuming the sequential three-state kinetics cannot be rationalized, and this is the reason we have ruled out the possibility of such kinetics.

Even if the stretched and the bifurcated kinetics cannot be distinguished in terms of SADSs, we favor the stretched exponential kinetics over the bifurcated kinetics based on the EOM analysis on the SAXS curve of the unfolded cyt-c as discussed in the main text. To check if the two kinetic models can be further discerned solely based on the experimental data, we adopted the kinetic analysis based on the maximum entropy method (MEM) whereby the time-dependent signal can be explained by a distribution of time constants (28). From this analysis, the optimal distribution of log-time constants can be extracted from the

experimental data as shown in Fig. S13. Prior to the application of this method to our experimental data, we implemented the benchmark analysis using two mock data with the same number of time points and the same time range as our experimental data: one mock data with a stretched exponential kinetics and the other mock data with a bifurcated kinetics. In the case of bifurcated kinetics, the relative ratio of two time constants determined from the bi-exponential fit for the experimental data was used. For better comparison, we used both the early and the late time domains for the MEM analysis shown in Fig. S13. In the case of stretched exponential kinetics, the optimal distributions extracted from the analyses for both of the early and the late time delays show the prominent populations in the region earlier than 10 ms. Contrary to this, the optimal distributions in the bifurcated kinetics show the extremely insignificant populations around 10 ms. The optimal distributions from the experimental data show similar features to those from the mock data with the stretched exponential kinetics. Therefore, we conclude that the MEM results also favor the stretched exponential kinetics over the bifurcated kinetics.

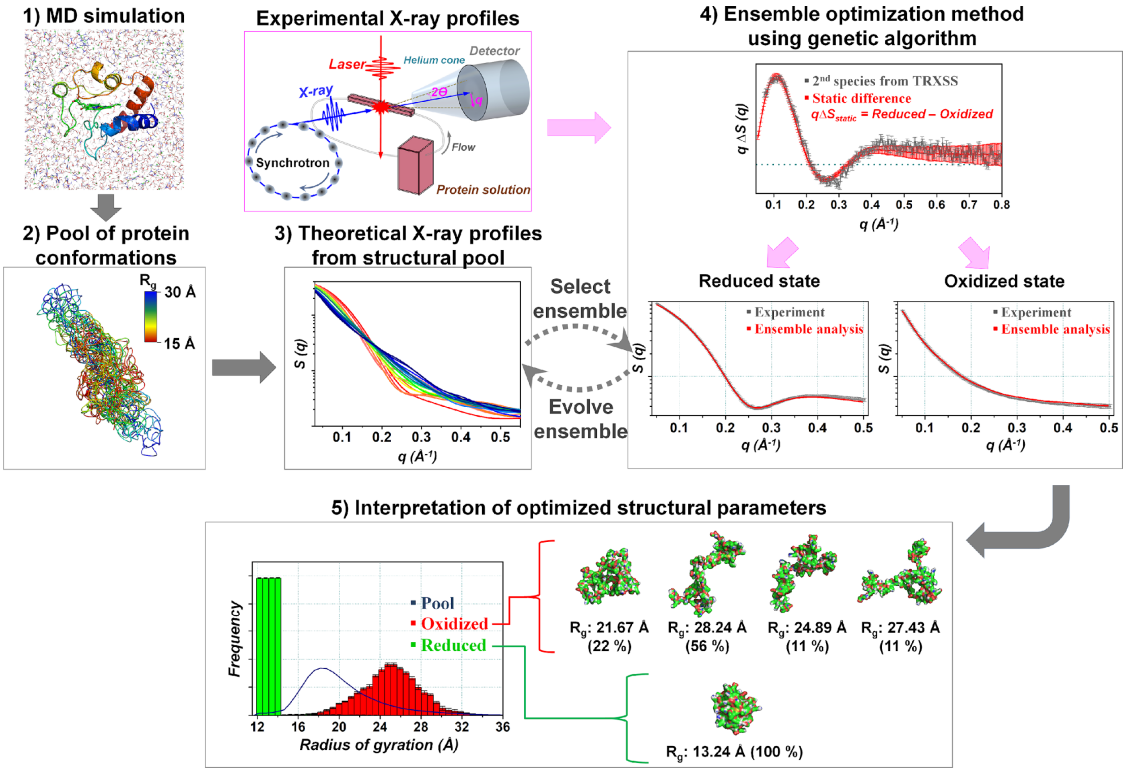

Fig. S1. Overview of the EOM based on the genetic algorithm to extract the structural parameters from the X-ray scattering data. The first step of this analysis is to make a pool of protein conformations from MD simulations and to generate the theoretical scattering curves that correspond to each protein conformation in the pool. An ensemble of some conformations are randomly selected from the pool and the composite X-ray scattering curve is obtained based on individual theoretical scattering curves. This is repeated for multiple ensembles and their computed X-ray scattering curves are compared with the experimental static X-ray data of the reduced and oxidized states. The protein structures from the best fitting ensembles are propagated into the next generation of ensembles along with some new structures introduced from the original conformational pool. This evolution process is reiterated until the discrepancy between the theoretical and the experimental scattering curves is minimized. The structural information for radius of gyration (R_g) and representative protein conformations with those fractions is determined from the final ensemble.

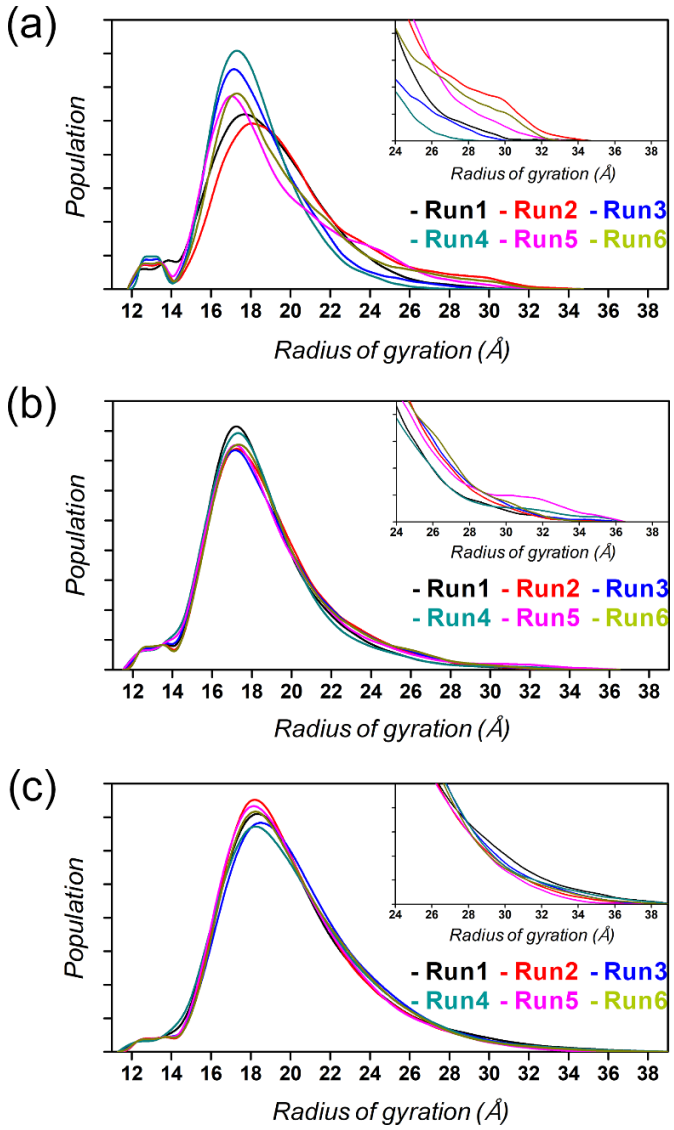

Fig. S2. Conformational space of protein structures as represented by R_g distributions sampled from MD simulations over (a) 100 ns, (b) 500 ns, and (c) 1 μs durations. As the simulation time increases, the maximum Rg value observed in the simulation increases from ~34 Å with 100 ns to ~38 Å with 1 μs. All the conformational spaces encompass the R_g values (24–29 Å) of unfolded cyt-c reported from previous experiments (15, 16).

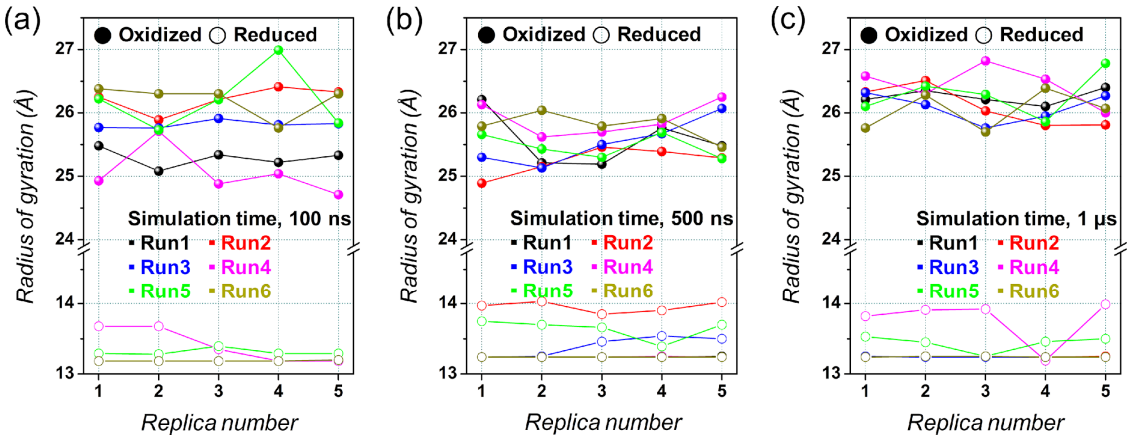

Fig. S3. Averaged radius of gyration of oxidized and reduced cyt-c determined from EOM. (a-c) Averaged radius of gyration (R_g) for the optimized ensemble of the oxidized state (solid circle) and the reduced state (open circle) determined from EOM based on MD simulations of protein unfolding with simulation durations of (a) 100 ns, (b) 500 ns and (c) 1 μs. To check the dependence of the result from EOM on the structure pool itself, we implemented the entire EOM analyses using six different MD runs as explained in the SI note. For each MD run, we repeated the EOM analysis five times to check the deviations of resulting R_g values. As the simulation time increases, the deviation of R_g values for the oxidized cyt-c decreases.

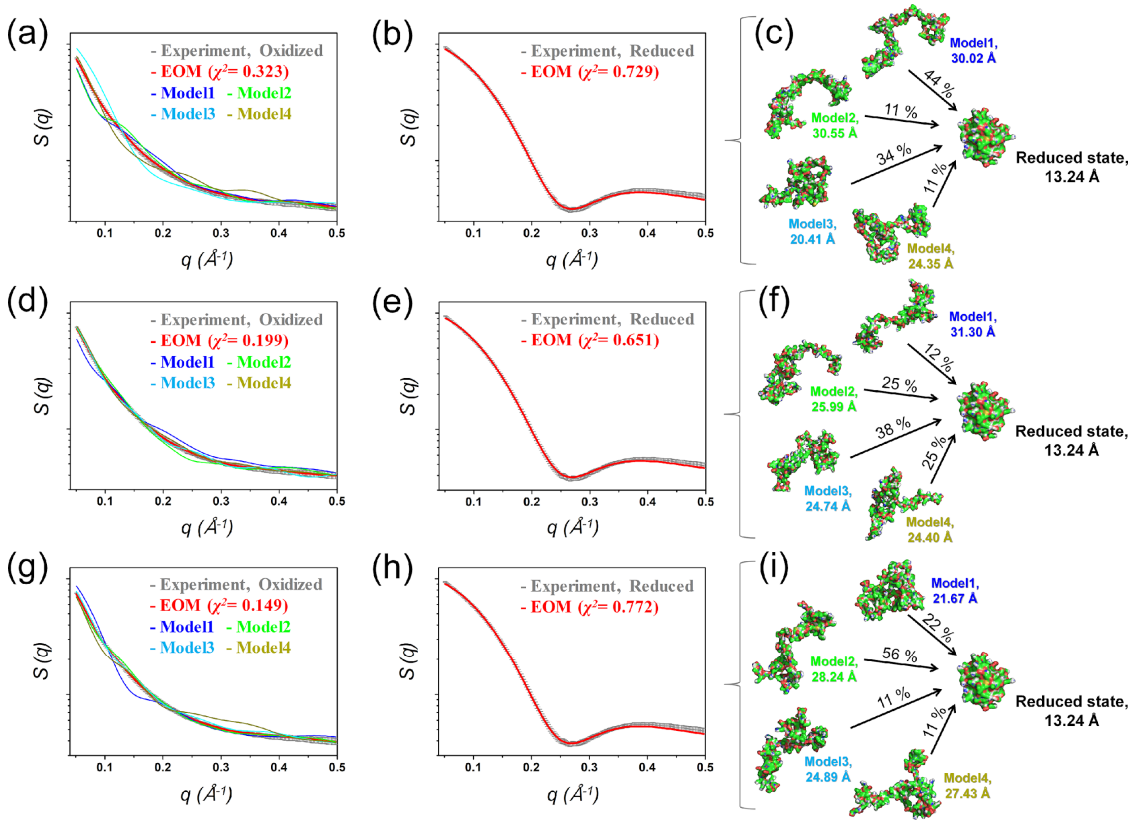

Fig. S4. Structural analysis of X-ray scattering curves based on the ensemble optimization method (EOM). (a) Based on one 100 ns MD trajectory, the X-ray scattering curve for the oxidized cyt-c was calculated from the final ensemble (red line) based on individual scattering curves by the representative structures in the ensemble (blue, green, light blue, yellow lines). This is compared with the experimental X-ray data of the unfolded state (gray line). The χ^2 value shows a good agreement between the theoretical scattering curve and the experimental one. (b) The same data as in (a), but for the reduced cyt-c. (c) Representative protein conformations with their fractions determined from EOM. (d-f) Results from EOM based on 500 ns simulations. (g-i) Results from EOM based on 1 μs simulations. In the optimized ensembles of the oxidized state, a single individual structure could not reproduce the experimental data. This aspect signifies that the oxidized state in 3.5 M GdnHCl has inherent conformational heterogeneity.

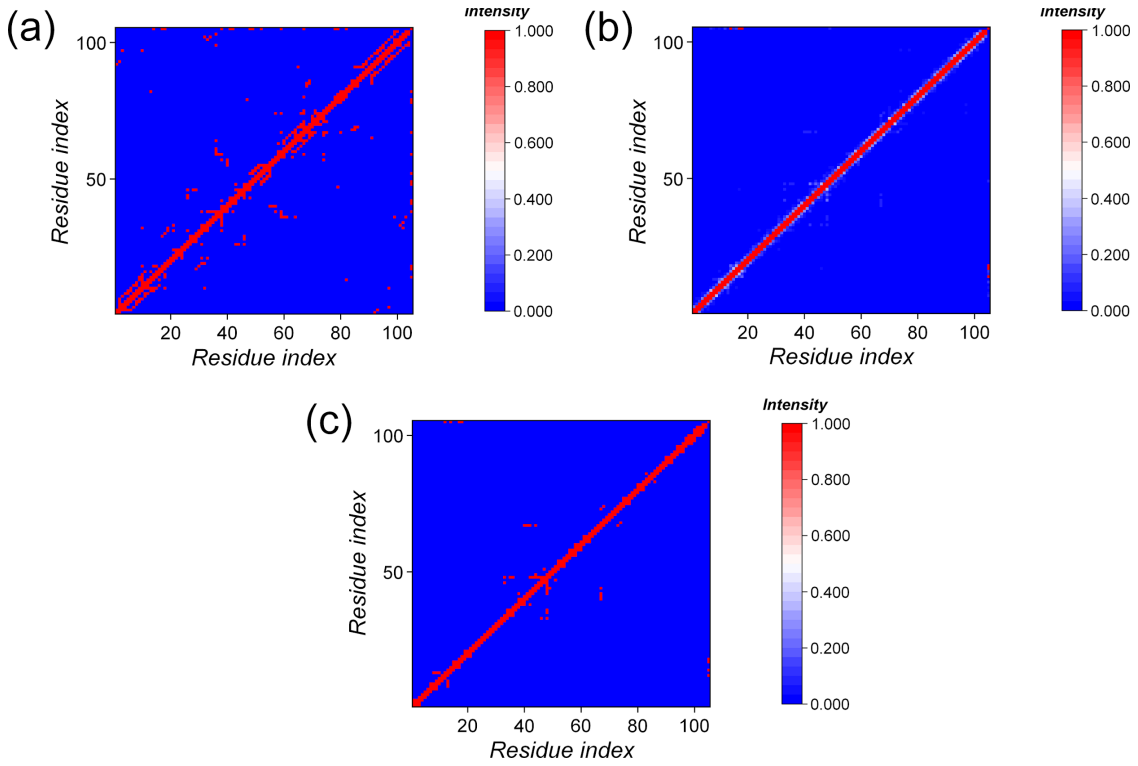

Fig. S5. Ensemble contact maps for (a) the folded state, (b) the unfolded state, and (c) a subset of the unfolded state. The subset in (c) was constructed by choosing unfolded structures with similar contact patterns and constitute ~7% of all structures in the unfolded ensemble. The contact maps were generated by averaging all individual contact maps from each ensemble, with the adjusted population values from EOM. A contact between a pair of residues was declared when any interatomic distance within that pair was shorter than 7 Å.

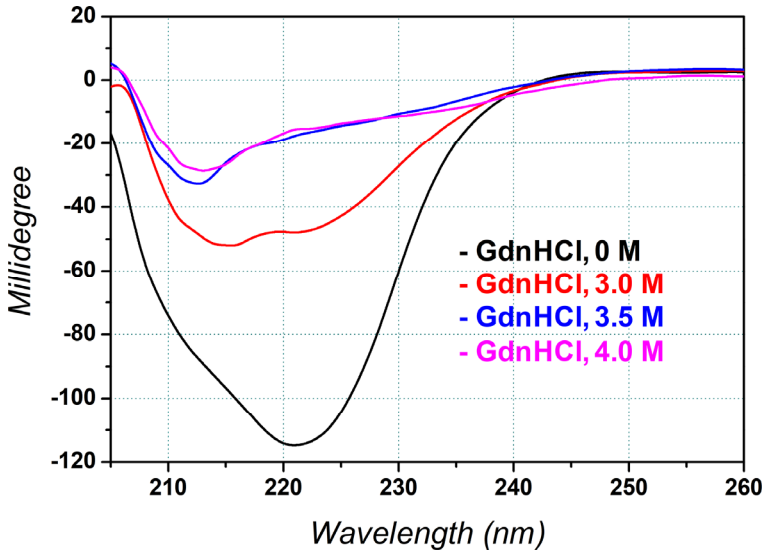

Fig S6. Far-UV CD spectra as a function of the concentration of guanidine hydrochloride, GdnHCl. The spectral shape of the native state prior to the addition of GdnHCl shows a large negative signal around 220 nm, indicating that the existence of alpha-helical structure in the folded state. As the concentration of GdnHCl increases, the negative shape around 220 nm is diminished. At 4.0 M GdnHCl, the CD spectrum shows a quite flat feature with the small negative intensity around 220 nm, implying that this denaturing condition causes the unfolding of cyt-c with a high content of random-coil structure. The spectrum with 3.5 M GdnHCl, used for the time-resolved X-ray scattering measurement, shows a similar feature with that with 4.0 M GdnHCl, supporting that the initial state in the condition of 3.5 M GdnHCl contains the fully unfolded state with random-coil conformations.

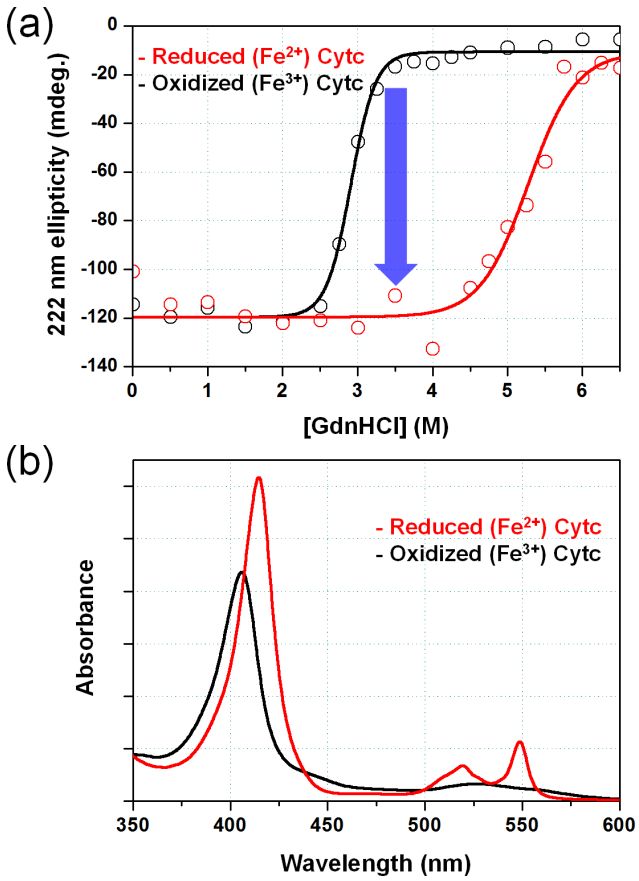

Fig. S7. Steady-state optical spectroscopic measurements. (a) The circular dichroism (CD) intensities of oxidized and reduced cyt-c at 222 nm as a function of the GdnHCl concentration dissolved in 100 mM sodium phosphate buffer solution. Cyt-c was reduced by adding sodium dithionite into the N_2-saturated protein solution. At the concentration of 3.5 M GdnHCl, the CD intensity shows a prominent decrease in the transition from the oxidized to the reduced form. This means that all secondary structure of reduced cyt-c is almost lost above the concentration of 3.5 M GdnHCl and the reduced cyt-c structure is more stable than the oxidized form. (b) UV-visible absorption spectra of oxidized and reduced cyt-c at 3.5 M GdnHCl.

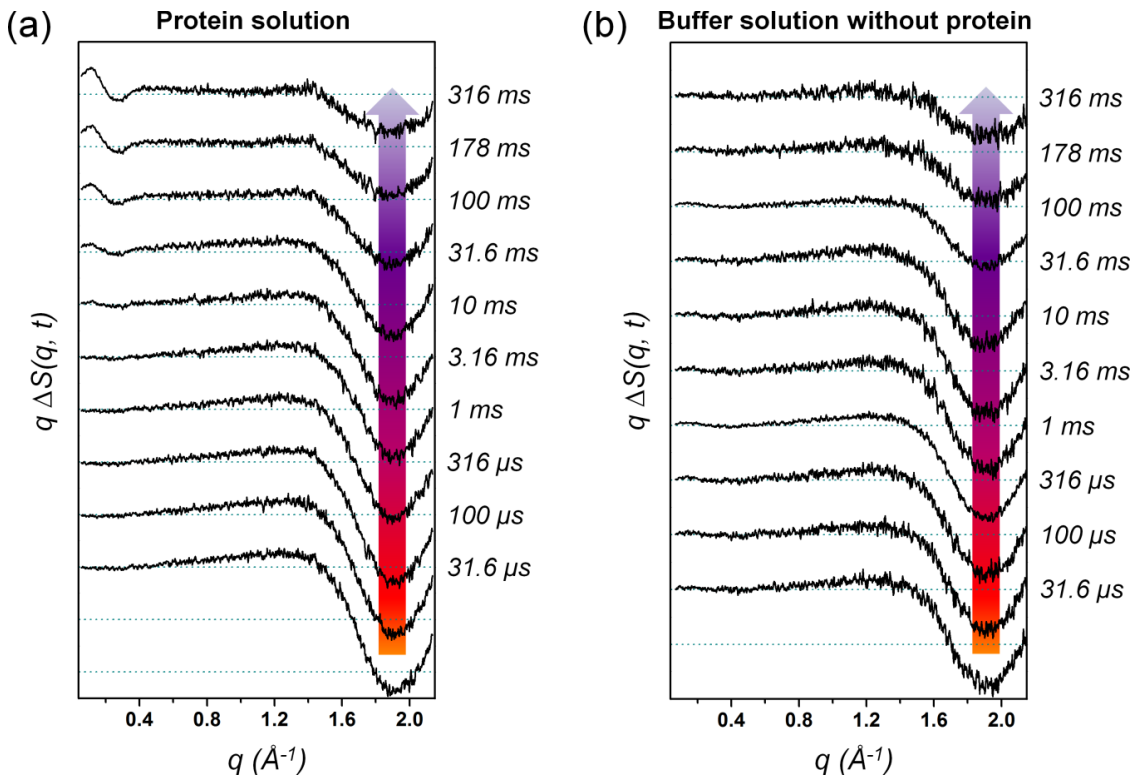

Fig. S8. Time-resolved X-ray solution scattering curves. (a) Time-resolved X-ray solution scattering curves for the protein solution. (b) TRXSS signal for the buffer solution without cyt-c. To check the effect of NADH used as an external electron donor in the time-resolved scattering signal, we performed a separate experiment for the buffer solution without cyt-c. In the small-angle region below 0.6 $Å^{-1}$, the neat buffer solution shows quite flat signal over the entire time window, whereas there is a typical heat response in the wide-angle region around 1.9 $Å^{-1}$. The colored arrows indicate the contribution of solvent heating induced by the dissipation of laser excitation energy.

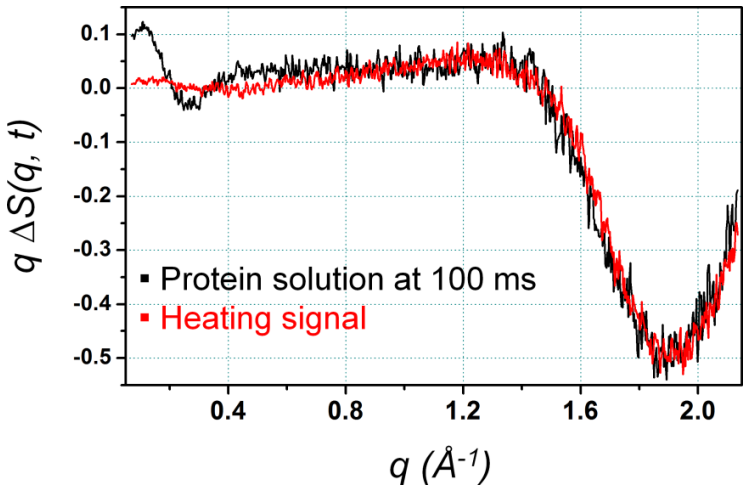

Fig. S9. Comparison of time-resolved X-ray solution scattering curve for protein solution and the heating signal. The solvent heating response from the pristine buffer solution (red line) is compared with the difference scattering curve of the protein solution at 100 ms time delay (black line). It can be seen that the small-angle region (q < 0.6 Å^{-1}) is not much affected by the solvent heating.

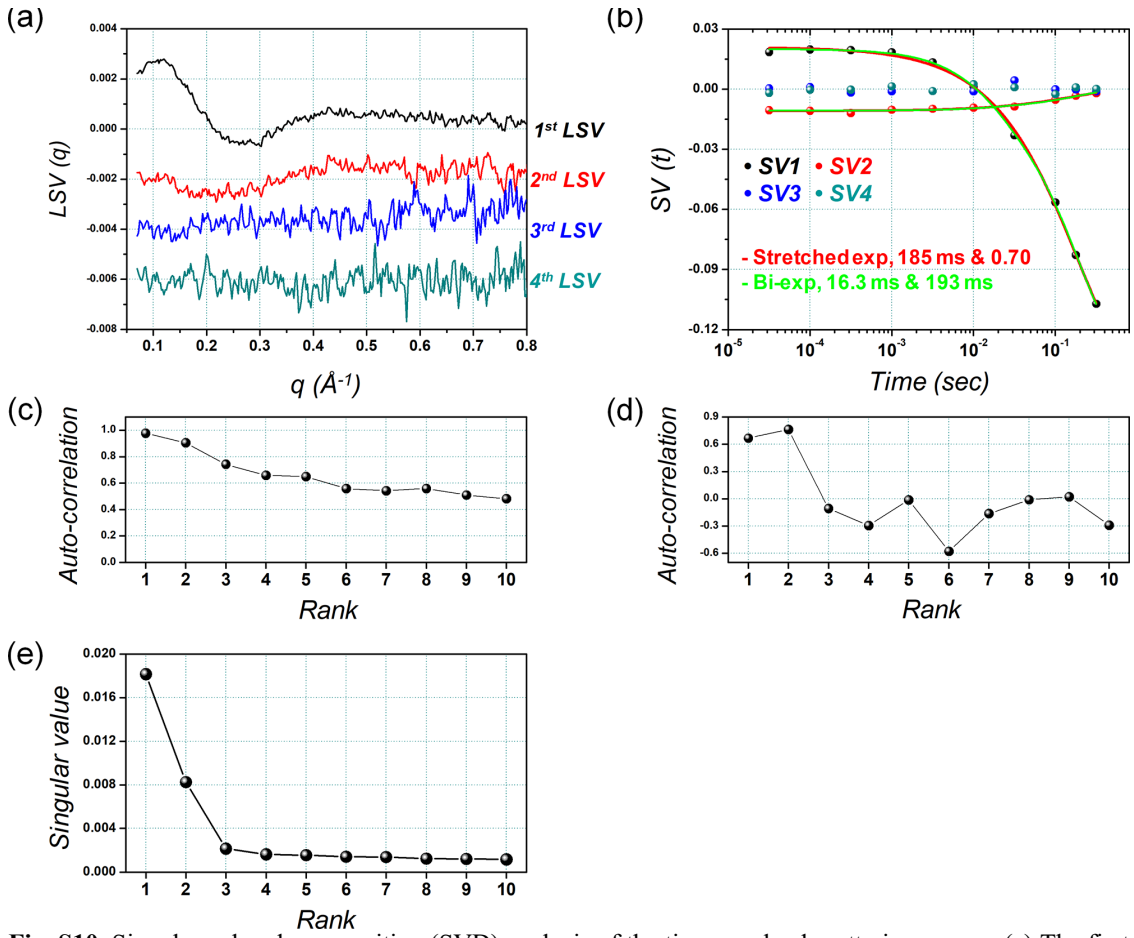

Fig. S10. Singular value decomposition (SVD) analysis of the time-resolved scattering curves. (a) The first four left singular vectors (LSVs). (b) The most significant right singular vectors (RSVs) multiplied by their corresponding singular values. The RSVs (dots) were globally fitted by a stretched exponential functions, $e^{-(t/\tau)^{\beta}}$ (red lines). The global fit for the RSVs using the stretched exponential functions yields the time constant (τ) of 185 ms and the β value of 0.70. For comparison, the bi-exponential fit with the time constants of 16.3 ms and 193 ms (green lines) is overlaid with the stretched exponential fit. Autocorrelation values of (c) LSVs and (d) RSVs. (e) Singular values. Considering the shape of singular vectors and their autocorrelation values, the first two LSVs and RSVs are significant for the data from 31.6 μs to 316 ms.

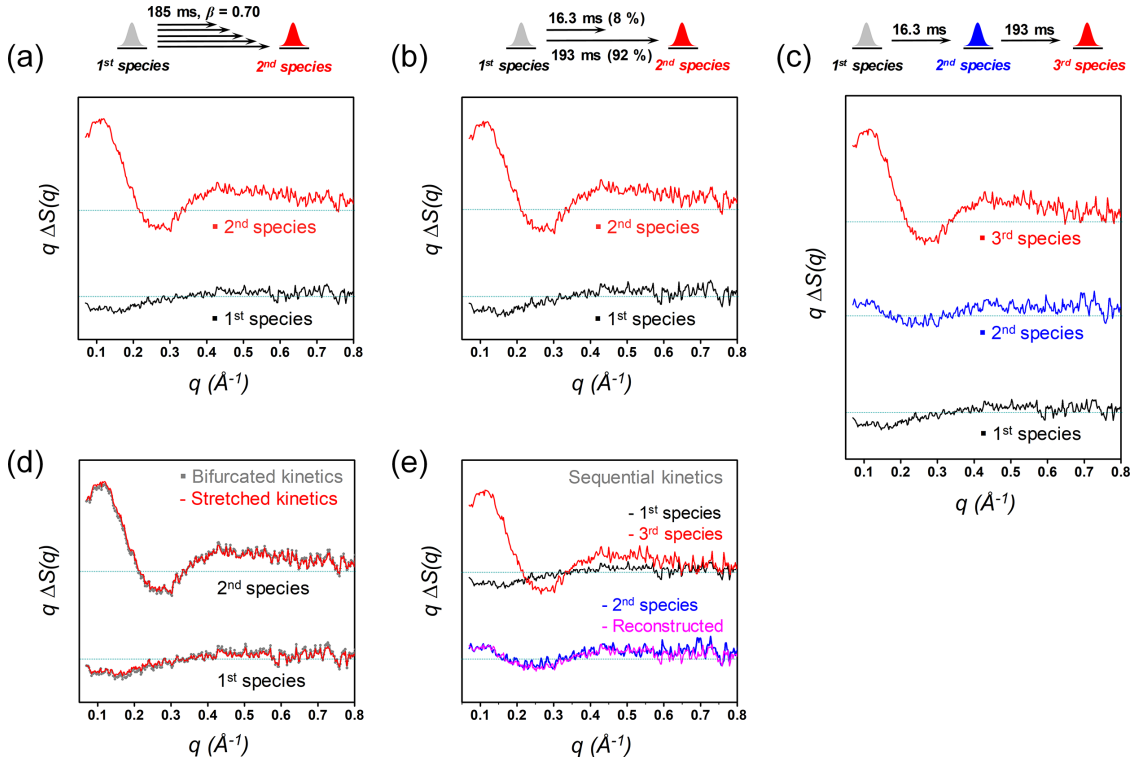

Fig. S11. Global kinetic analysis based on possible candidate models. (a-c) SADSs extracted from (a) the stretched exponential kinetics with the time constant of 185 ms, (b) the bifurcated kinetics accompanying the parallel pathways with two time constants of 16.3 ms and 193 ms, and (c) the sequential kinetics with the consecutive decays with two time constants of 16.3 ms and 193 ms. (d) Comparison of SADSs from the stretched exponential kinetics and those from the parallel kinetics. The good agreement shows that two kinetic models are indistinguishable in terms of the spectral shape of SADSs. (e) Verification of the sequential kinetic model. The reconstructed difference scattering curve from the linear combination of 1st and 3rd SADS is in good agreement with the 2nd SADS, showing that the 2nd SADS is not essential to describe the experimental data.

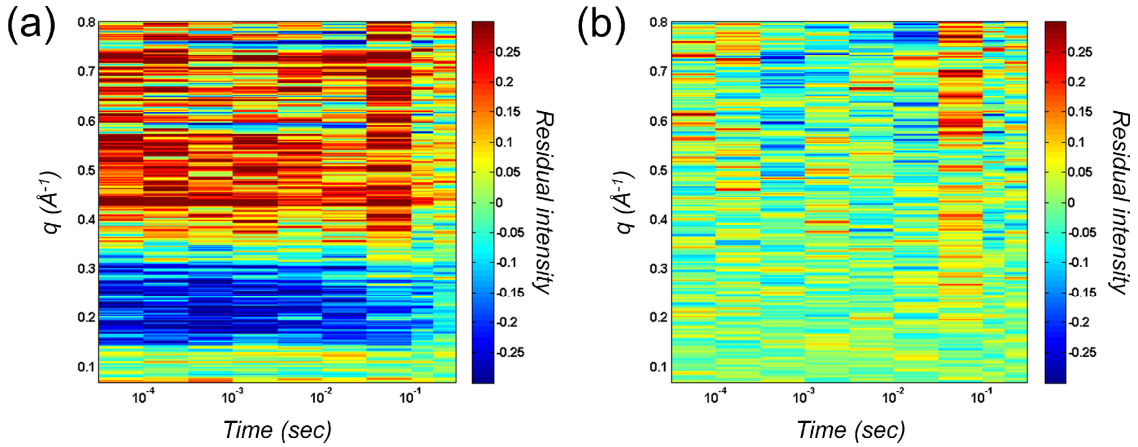

Fig. S12. Residual map between the experimental data and the fitted models, (a) with the fit based on only the second species in the folding dynamics and (b) with the fit using both the first and the second species in the folding dynamics. It is obvious that the omission of the first species in the folding kinetics results in a much worse agreement.

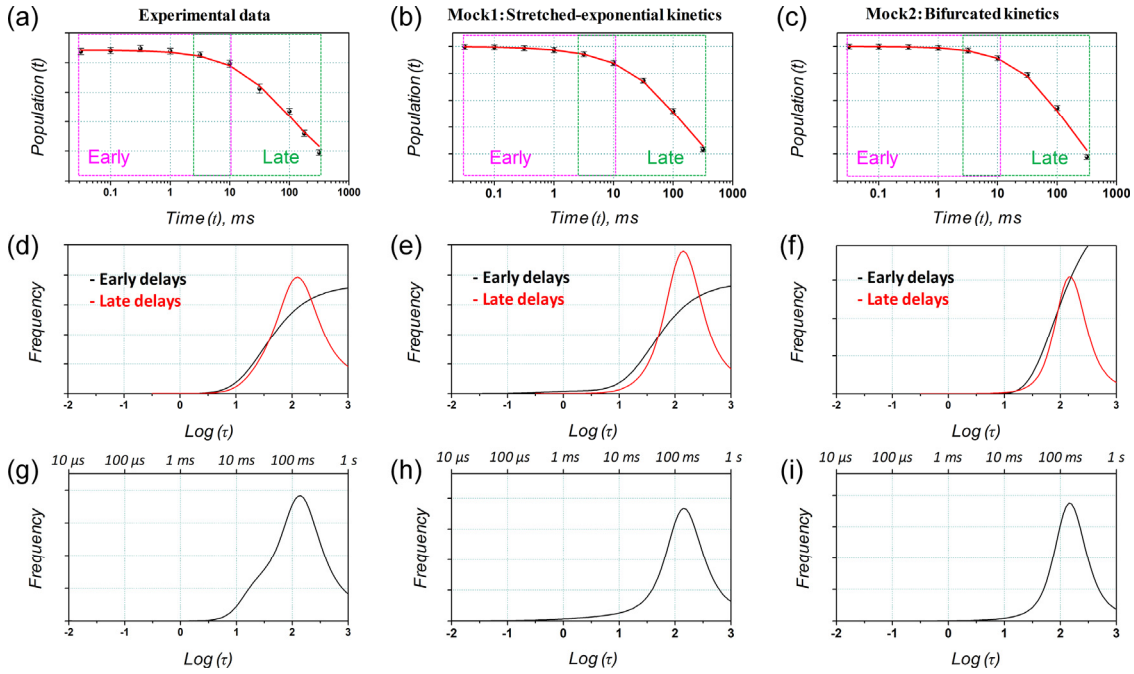

Fig. S13. Maximum entropy method (MEM) analyses of the experimental observed and the theoretically modelled data (mock data). (a) Analysis of the first right singular vector. The fitted result (red line) with the χ^2 value of 0.982 shows a good agreement with the experimental date (black dots). (b) Analysis of the mock data with a stretched exponential kinetics with the time constant of 185 ms and the β value of 0.70. (c) Analysis with bifurcated kinetics with two time constants of 16.3 and 193 ms. In bifurcated kinetics, the relative ratio (pre-exponential factor) of two time constants determined from the bi-exponential fit for the experimental data was used. For better comparison, in addition to the entire time range (31.6 μs – 316 ms), we also used the early (31.6 μs – 10 ms) and the late (3.16 ms – 316 ms) time domains for the MEM analysis. (d-f) Optimized distributions from (d) the experimental data, (e) the mock data with a stretched exponential kinetics, and (f) the mock data with a bifurcated kinetics determined from the MEM analyses using the limited time windows. The red and the black curves are from the early and the late time windows, respectively. (g-i) Optimized distributions from the MEM analyses for the entire time range in (g) the experimental data, (h) the mock data with the stretched exponential kinetics and (e) the mock data with the bifurcated kinetics. The zero value in the x-axis corresponds to 1 ms.

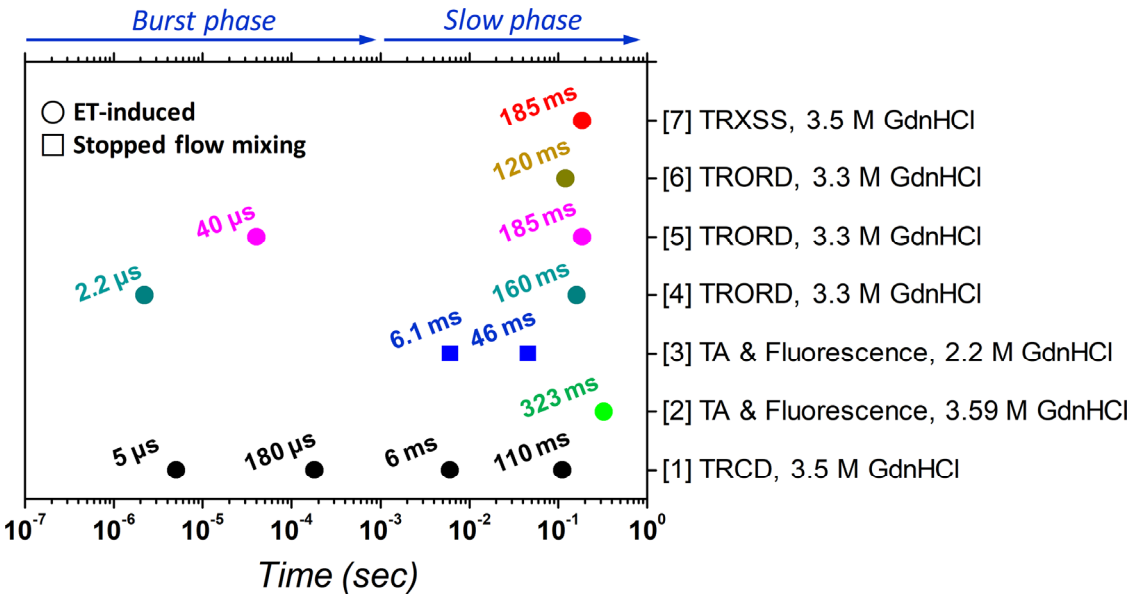

Fig. S14. Comparison of time constants determined from earlier studies of cyt-c folding based on electron-transfer induced (ET-induced) and stopped flow mixing methods. The labels on the right show the details about the probing method for monitoring the protein folding together with the adopted GdnHCl concentrations. The works based on T-jump, pH-jump, or CO-photolysis are not included here. Details of the labels are as follows.

[1] Time-resolved circular dichroism spectroscopy (29); [2] Transient absorption & fluorescence spectroscopy (30); [3] Transient absorption & fluorescence spectroscopy (31); [4] Time-resolved optical rotatory dispersion spectroscopy (32); [5] Time-resolved optical rotatory dispersion spectroscopy for tuna heart cyt-c (33); [6] Time-resolved optical rotatory dispersion spectroscopy for H26QH33N mutant cyt-c (34); and [7] the current study using time-resolved X-ray solution scattering.

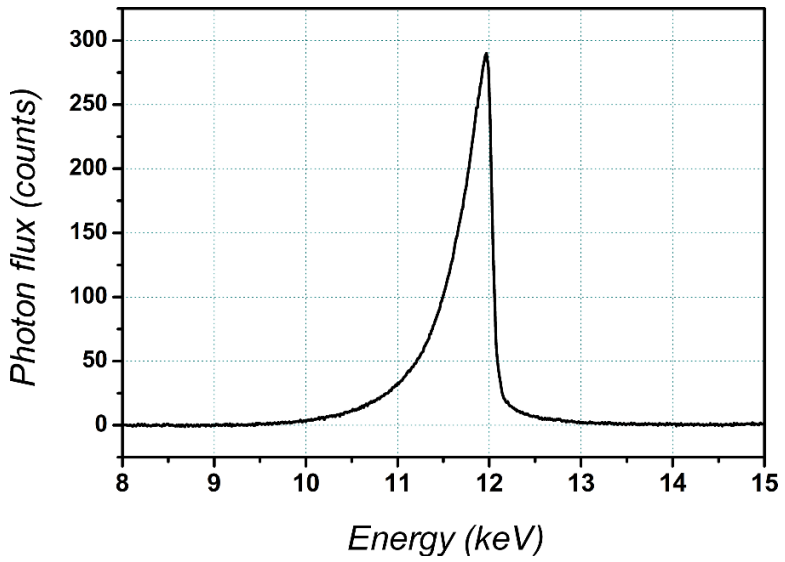

Fig. S15. Energy spectrum of an X-ray pulse with a characteristic half-Gaussian shape, generated from the combination of U23 and U27 undulators in the 14-ID-B beamline at APS. All the theoretical scattering curves used in the EOM were convoluted with the measured energy spectrum in order to correct the effect of polychromatic X-ray spectrum on the scattering curve.

References

1. Bushnell GW, Louie GV, & Brayer GD (1990) High-Resolution 3-Dimensional Structure of Horse Heart Cytochrome-C. *J. Mol. Biol.* 214(2):585-595.
2. Van Der Spoel D*, et al.* (2005) GROMACS: fast, flexible, and free. *J. Comput. Chem.* 26(16):1701-1718.
3. Daura X, Mark AE, & van Gunsteren WF (1998) Parametrization of aliphatic CHn united atoms of GROMOS96 force field. *J. Comput. Chem.* 19(5):535-547.
4. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, & Klein ML (1983) Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* 79(2):926-935.
5. Bussi G, Donadio D, & Parrinello M (2007) Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126(1).
6. Camilloni C*, et al.* (2008) Urea and guanidinium chloride denature protein L in different ways in molecular dynamics simulations. *Biophys. J.* 94(12):4654-4661.
7. Darden T, York D, & Pedersen L (1993) Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* 98(12):10089-10092.
8. Essmann U*, et al.* (1995) A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* 103(19):8577-8593.
9. Tolman RC (1925) The Principle of Microscopic Reversibility. *Proc. Natl. Acad. Sci. U.S.A.* 11(7):436-439.
10. Toofanny RD & Daggett V (2012) Understanding protein unfolding from molecular simulations. *Comput. Mol. Sci.* 2(3):405-423.
11. Wang T & Wade RC (2007) On the use of elevated temperature in simulations to study protein unfolding mechanisms. *J. Chem. Theory Comput.* 3(4):1476-1483.
12. Qiu D, Shenkin PS, Hollinger FP, & Still WC (1997) The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii. *J. Phys. Chem. A* 101(16):3005-3014.
13. Elove GA, Bhuyan AK, & Roder H (1994) Kinetic mechanism of cytochrome c folding: involvement of the heme and its ligands. *Biochemistry* 33(22):6925-6935.
14. Colon W, Wakem LP, Sherman F, & Roder H (1997) Identification of the predominant non-native histidine ligand in unfolded cytochrome c. *Biochemistry* 36(41):12535-12541.
15. Akiyama S*, et al.* (2002) Conformational landscape of cytochrome c folding studied by microsecond-resolved small-angle X-ray scattering. *Proc. Natl. Acad. Sci. U.S.A.* 99(3):1329-1334.
16. Segel DJ, Fink AL, Hodgson KO, & Doniach S (1998) Protein denaturation: A small-angle X-ray scattering study of the ensemble of unfolded states of cytochrome c. *Biochemistry* 37(36):12443-12451.
17. Svergun D, Barberato C, & Koch MHJ (1995) CRYSOL - A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* 28:768-773.
18. Kim KH, Kim J, Lee JH, & Ihee H (2014) Topical Review: Molecular reaction and solvation visualized by time-resolved X-ray solution scattering: Structure, dynamics, and their solvent dependence. *Struct. Dynam.* 1(1):011301.
19. Kim TK, Lee JH, Wulff M, Kong QY, & Ihee H (2009) Spatiotemporal Kinetics in Solution Studied by Time-Resolved X-Ray Liquidography (Solution Scattering). *ChemPhysChem* 10(12):1958-1980.
20. Jones G (1998) Genetic and Evolutionary Algorithms. (Wiley, Chichester, UK):pp 1127-1136.
21. Bernado P, Mylonas E, Petoukhov MV, Blackledge M, & Svergun DI (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. *J. Am. Chem. Soc.* 129(17):5656-5664.
22. Kim TW*, et al.* (2016) Combined probes of X-ray scattering and optical spectroscopy reveal how global conformational change is temporally and spatially linked to local structural perturbation in photoactive yellow protein. *Phys. Chem. Chem. Phys.* 18(13):8911-8919.
23. Kim TW*, et al.* (2012) Protein Structural Dynamics of Photoactive Yellow Protein in Solution Revealed by Pump-Probe X-ray Solution Scattering. *J. Am. Chem. Soc.* 134:3145-3153.
24. Kim KH*, et al.* (2012) Direct Observation of Cooperative Protein Structural Dynamics of Homodimeric Hemoglobin from 100 ps to 10 ms with Pump-Probe X-ray Solution Scattering. *J. Am. Chem. Soc.* 134(16):7001-7008.

25. Cammarata M*, et al.* (2006) Impulsive solvent heating probed by picosecond X-ray diffraction. *J. Chem. Phys.* 124(12).
26. Oang KY, Yang C, Muniyappan S, Kim J, & Ihee H (2017) SVD-aided pseudo principal-component analysis: A new method to speed up and improve determination of the optimum kinetic model from time-resolved data. *Struct. Dynam.* 4(4):044013.
27. Lagarias JC, Reeds JA, Wright MH, & Wright PE (1998) Convergence properties of the Nelder-Mead simplex method in low dimensions. *Siam. J. Optimiz.* 9(1):112-147.
28. Kumar ATN, Zhu L, Christian JF, Demidov AA, & Champion PM (2001) On the Rate Distribution Analysis of Kinetic Data Using the Maximum Entropy Method: Applications to Myoglobin Relaxation on the Nanosecond and Femtosecond Timescales. *J. Phys. Chem. B* 105(32):7847-7856.
29. Chen EF, Wittung-Stafshede P, & Kliger DS (1999) Far-UV time-resolved circular dichroism detection of electron-transfer-triggered cytochrome c folding. *J. Am. Chem. Soc.* 121(16):3811-3817.
30. Pascher T (2001) Temperature and driving force dependence of the folding rate of reduced horse heart cytochrome c. *Biochemistry* 40(19):5812-5820.
31. Bhuyan AK & Udgaonkar JB (2001) Folding of horse cytochrome c in the reduced state. *J. Mol. Biol.* 312(5):1135-1160.
32. Chen EF, Goldbeck RA, & Kliger DS (2003) Earliest events in protein folding: Submicrosecond secondary structure formation in reduced cytochrome c. *J. Phys. Chem. A* 107(40):8149-8155.
33. Chen E, Goldbeck RA, & Kliger DS (2004) The earliest events in protein folding: a structural requirement for ultrafast folding in cytochrome C. *J. Am. Chem. Soc.* 126(36):11175-11181.
34. Chen E, Abel CJ, Goldbeck RA, & Kliger DS (2007) Non-native heme-histidine ligation promotes microsecond time scale secondary structure formation in reduced horse heart cytochrome c. *Biochemistry* 46(43):12463-12472.